

End-to-End Model Methodology

Nick Singer

21 February 1996



Workload Characterization

Threads in Place

- Inventory Search (Motif client) 2880/day
- Retrieve (Browse) 1200/day
- Acquire via Network 1620/day
- Acquire via Media 1620/day
- Insert production result to DS 4003/day
- Insert L0 to Ingest 329/day

Threads Ready to Go In Now

- Inventory Search (Web client)
- Browse Search (Motif client)
- Browse Search (Web client)

Threads by End of Month

- Log-on
- Document Data Server queries and updates
- Planning and Scheduling activities
- Maintenance-driven activities (e.g. data server monitoring services)

Workload Characterization (cont'd)

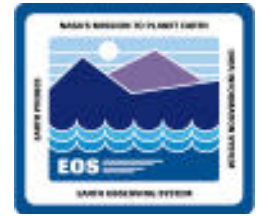


Threads

- Instantiation rate (per second)
- List of activities

Activities

- Specific CPU
 - CPU Instructions (millions)
 - Custom code
 - RPCs
 - Distributed object instantiations
 - DBMS calls
- Network(s)
- # Network transfers
- Avg. MB/network transfer
- Disk
- Avg. MB/disk transfer

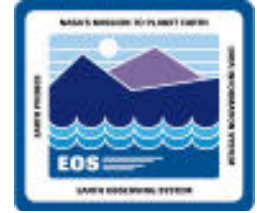


System Characterization



Hardware

- **Machines (CPUs)**
 - MIPS
- **Networks**
 - Sustainable throughput rate (MB/sec)
 - Switch latency time
- **Disks**
 - Sustainable transfer rate (MB/sec)
 - Latency time



Methodology

Read characterization files; set up model

Step through threads & activities; collect statistics on load by thread and in total

- By specific CPU
- By specific network
- By specific disk

Add in known background loads for each resource

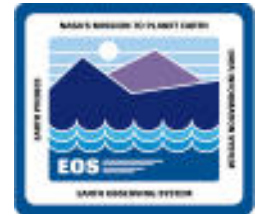
Calculate utilization & expected waiting time at each resource

- From above statistics, can calculate mean and variance of service time at each node
- Use Pollaczek-Khinchin mean-value formula for waiting time (M/G/1)

$$W = \frac{\lambda E[\text{service time}^2]}{2(1 - \rho)}$$

Calculate end-to-end times for each thread

- Time for a given activity at a given resource = average waiting time for the resource + service time for the activity at the resource



End-to-End Model Next Steps



Complete the analysis and integration of remaining threads and activities

Integrate

- **Infrastructure analysis & results**
- **Dynamic modeling results**

Make model runs at nominal and higher load levels